Journal of Computational and Applied Mathematics 233 (2010) 2683–2687

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Improving weighted information criterion by using optimization

Cagdas Hakan Aladag [a,*], Erol Egrioglu [b], Suleyman Gunay [a], Murat A. Basaran [c]

[a] Department of Statistics, Hacettepe University, Ankara 06800, Turkey

[b] Department of Statistics, Ondokuz Mayıs University, Samsun 55100, Turkey

[c] Department of Mathematics, Nigde University, Nigde 51000, Turkey

ARTICLE INFO

Article history:
Received 29 June 2009
Received in revised form 11 November 2009

Keywords:
Artificial neural networks
Consistency
Forecasting
Model selection
Time series
Weighted information criterion

ABSTRACT

Although artificial neural networks (ANN) have been widely used in forecasting time series, the determination of the best model is still a problem that has been studied a lot. Various approaches available in the literature have been proposed in order to select the best model for forecasting in ANN in recent years. One of these approaches is to use a model selection strategy based on the weighted information criterion (WIC). WIC is calculated by summing weighted different selection criteria which measure the forecasting accuracy of an ANN model in different ways. In the calculation of WIC, the weights of different selection criteria are determined heuristically. In this study, these weights are calculated by using optimization in order to obtain a more consistent criterion. Four real time series are analyzed in order to show the efficiency of the improved WIC. When the weights are determined based on the optimization, it is obviously seen that the improved WIC produces better results.

1. Introduction

ANN is a method which has been successfully used in many areas for different purposes. Although ANN has proved its ability in various applications, the research for improving this method is still under way. Determining the elements of ANN is a vital subject in order to improve the method. In order to determine the best elements of ANN, there have been some used approaches such as polynomial time algorithm [1], canonical decomposition [2], network information criterion [3], iterative construction algorithm [4], pruning algorithm [5–7], a strategy based on Box–Jenkins method [8], a method based on information entropy [9], genetic algorithm [10], principal component analysis [11], weighted information criterion proposed in [12], and deletion/substitution/addition algorithm [13].

Due to the ability of modeling both linear and non-linear structures, ANN has been used widely in the forecasting of time series in many areas. Determining the elements of ANN is an important factor for obtaining high accuracy forecasts. Some criteria are used to determine the best ANN model in forecasting. Egrioglu et al. proposed *WIC* in order to find the best ANN model [12]. It was shown that *WIC* is a more consistent criterion than the root mean square error (*RMSE*) criterion, which has been generally used in the literature. Weighted information criterion consists of summing the weighted different criteria. Egrioglu et al. determined these weights intuitively [12]. In this paper, we improved the *WIC* criteria by optimizing the weights used as coefficients in the *WIC* criterion. This improvement provides three advantages. One of them is that the weights are determined systematically. Second, the weights are calculated using the available data. That is, the weights change according to the analyzed data. The third is that the improved *WIC* becomes more consistent. The newly defined *WIC* is called the adaptive weighted information criteria (*AWIC*) since its weights are determined with respect to the data

* Corresponding author.
E-mail address: chaladag@gmail.com (C.H. Aladag).

0377-0427/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2009.11.016

which are analyzed. *AWIC* is applied to four real time series data in order to measure forecasting performance of examined feed forward ANN architectures and the obtained results are compared with those from *WIC*. It is clearly seen that *AWIC* is more consistent than *WIC*.

Section 2 contains in the model selection criterion based on *WIC* proposed in [12]. Section 3 includes a systematic approach in order to determine weights using optimization. Four real time series are analyzed by using *AWIC* and obtained results are given in Section 4. Paper is concluded in Section 5.

2. The model selection strategy based on *WIC*

WIC criterion consists of summing weighted *AIC*, *BIC*, *RMSE*, *MAPE*, *DA*, and *MDA* criteria. The expressions related to the criteria are given below.

$$AIC = \log\left(\frac{\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}{T}\right) + \frac{2m}{T} \tag{2.1}$$

$$BIC = \log\left(\frac{\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}{T}\right) + \frac{m\log(T)}{T} \tag{2.2}$$

$$RMSE = \left(\frac{\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}{T}\right)^{1/2} \tag{2.3}$$

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2.4}$$

$$DA = \frac{1}{T}\sum_{i=1}^{T} a_i, \quad a_i = \begin{cases} 1 & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

where y_i is the actual value; $\hat{y}_i$ is the predicted value; T is the number of data and m is the number of weights. And *MDA* criterion is computed as follows.

$$\begin{aligned} &A_i = 1, \quad y_{i+1} - y_i \le 0 \\ &A_i = 0, \quad y_{i+1} - y_i > 0 \\ &F_i = 1, \quad \hat{y}_{i+1} - \hat{y}_i \le 0 \\ &F_i = 0, \quad \hat{y}_{i+1} - \hat{y}_i > 0 \\ &D_i = (A_i - F_i)^2 \\ &MDA = \frac{\sum_{i=1}^{T-1} D_i}{T-1}. \end{aligned} \tag{2.6}$$

The algorithm of the model selection strategy based on *WIC* is introduced below.

Step 1: Possible architectures are determined. For example, number of nodes of output layer is 1, number of nodes of input layer is 12 and number of nodes of hidden layer is 12. In this case, totally possible architecture number is 144.

Step 2: Best values of weights are determined by using training data and *AIC*, *BIC*, *RMSE*, *MAPE*, *DA* and *MDA* are calculated for test data.

Step 3: *AIC*, *BIC*, *RMSE*, *MAPE*, *DA* and *MDA* are standardized for possible architecture. For example 144 *AIC* values are standardized as follows:

$$AIC_i = \frac{AIC_i - \min(AIC)}{\max(AIC) - \min(AIC)}.$$

Step 4: *WIC* is computed in the following way.

$$WIC = 0.1(AIC + BIC) + 0.2(RMSE + MAPE) + 0.2\left((1 - DA) + MDA\right). \tag{2.7}$$

Step 5: Architecture, which has minimum *WIC*, is selected.

Egrioglu et al. [12] assigned the weights intuitively in (2.7). Since *AIC* and *BIC* often select the smallest models, their weights are determined as 0.1 by Egrioglu et al. The authors considered that the other criteria have same importance for forecasting accuracy so their weights are determined as 0.2.

3. Determining the weights of the *WIC* criterion by using optimization

In this study, *WIC* is improved by determining the weights in the criterion by using optimization. In order to make *WIC* more consistent in the selection of ANN models, the coefficients of *WIC* that maximize the correlation between test sets are computed. Time series is split into two sets for training and test. Then, test set divided into two equal subsets. For example, a time series containing 100 observations are split into training and test set, including 80 and 20 observations, respectively. Then, test set is partitioned into two equal subsets which consist of 10 observations. *RMSE*, *MAPE*, *DA*, *MDA*, *AIC* and, *BIC* which are obtained from different ANN models, are calculated for each two test sets. These criteria values are standardized by mapping into [0, 1] interval.

For the first and second test sets, $AWIC_1(i)$ and $AWIC_2(i)$ are calculated using the expressions below.

$$\begin{aligned} AWIC_1(i) &= w_1 RMSE_1(i) + w_2 MAPE_1(i) + w_3(1 - DA_1(i)) + w_4 MDA_1(i) + 0.1AIC_1(i) \\ &\quad + 0.1BIC_1(i), \quad i = 1, 2, \ldots, m_i \times m_h \\ AWIC_2(i) &= w_1 RMSE_2(i) + w_2 MAPE_2(i) + w_3(1 - DA_2(i)) + w_4 MDA_2(i) + 0.1AIC_2(i) \\ &\quad + 0.1BIC_2(i), \quad i = 1, 2, \ldots, m_i \times m_h \end{aligned}$$

where w_j $(j = 1, 2, 3, 4)$ denotes the weights or coefficients, m_i and m_h denote the number of neurons in input and hidden layer, respectively. The maximum number of architectures which are examined is obtained by multiplying m_i by m_h since output layer has just one neuron. *i* index used in the expressions above denotes architectures number. For example, $AWIC_1(i)$ is the value of *AWIC* which is calculated based on the first test set using the *i*th architecture and $MDA_2(i)$ is the value of *MDA* which is calculated based on the second test set using the *i*th architecture.

The coefficients of *BIC* and *AIC* are taken as fixed values which are 0.1. Since *AIC* and *BIC* criteria are dependent upon the number of the neurons in the investigated architecture, their consistency are high. This can be easily seen in the formulas of *AIC* and *BIC* given in (2.1) and (2.2), respectively. In these formulas, if the magnitude of *m* becomes greater, then the values of those criteria also grow greater in magnitude. The value of *m* will increase if the number of neurons increases. Thus, the more the number of neurons are, the higher the values of *AIC* and *BIC* become. On the other hand, other criteria called *RMSE*, *MAPE*, *DA* and *MDA* are not severely affected by the number of neurons. Therefore, *AIC* and *BIC* are more consistent criteria when they are compared with other criteria used in *WIC*. When *AIC* and *BIC* are employed to determine the best architecture, generally small architectures will be selected as the best architecture since those always have smaller *AIC* and *BIC* values than those with larger architectures. Thus, the coefficients of *AIC* and *BIC* should be taken as fixed values in the optimization process of the coefficients in *WIC*. Unless the weights of these two criteria are kept fixed, the weights of the other criteria converge to zero in the optimization process. Then, the coefficients of *BIC* and *AIC* are taken as 0.1 in the same way done in [12].

The weights which enable to reach maximum correlation value r_1 between the values of $AWIC_1(i)$ and $AWIC_2(i)$ are calculated. All initial values for w_1, w_2, w_3 and, w_4 are set to 0.2 at the beginning of the optimization process since Egrioglu et al. used the same values for the weights in *WIC* [12]. Thus, the obtained *AWIC* will be more consistent criterion than *WIC* whose weights are taken fixed. All calculations are done by using MATLAB 7.0 version.

4. Application

In order to show the consistency of *AWIC*, the test set is partitioned into three equal subsets in the applications of four real time series. The first and second test sets are used in order to determine the weights of *AWIC*. The way of determining the weights is explained in Section 3 of this paper. The third test is used to denote the consistency of the calculated *AWIC*. For the third test set, the values of $AWIC_3(i)$ are calculated by using the weights that are determined in the previous optimization step utilizing the first and second test sets. The formula is given below.

$$\begin{aligned} AWIC_3(i) &= w_1 RMSE_3(i) + w_2 MAPE_3(i) + w_3(1 - DA_3(i)) + w_4 MDA_3(i) + 0.1AIC_3(i) \\ &\quad + 0.1BIC_3(i), \quad i = 1, 2, \ldots, m_i \times m_h. \end{aligned}$$

The correlation value r_2 between the values of $AWIC_2(i)$ and $AWIC_3(i)$ is calculated. r_1 and r_2 correlation coefficients are compared with each other using four real time series in order to show the consistency of *AWIC*.

Four real time series are analyzed in the investigation of *AWIC* criterion. Both *AWIC* and *WIC* criteria are used to select the best ANN models for each time series in order to reach high forecasting accuracy. Proportion of imports covered by exports (PICE) and share in GNP of export (SE) time series are used in the implementation. These series consist of 81 observations were also used in [12]. In addition, monetary values of imports (MVI) and measures of sulfur dioxide (SO_2) in the air for Ankara (ANSO) time series are analyzed and these series have 72 and 92 observations, respectively. The ratio for training and test set is taken 0.15. For example, training and test set have 69 and 12 observations, respectively for PICE. When test set is divided into three test sets, each of them has 4 observations.

Table 1
The results for *WIC* and *AWIC* criteria.

Data	*WIC*			*AWIC*		
	r_1	r_2	Best architecture	r_1	r_2	Best architecture
PICE	0.6039	0.6871	2-11-1	0.9500	0.9397	3-1-1
SE	0.2468	0.3449	10-6-1	0.4400	0.6689	1-3-1
MVI	0.6515	0.7620	2-2-1	0.7507	0.8922	2-2-1
ANSO	0.5455	0.5951	6-2-1	0.7112	0.7112	4-5-1

Table 2
The weights of *AWIC*.

Data	w_1	w_2	w_3	w_4	w_5	w_6
PICE	0.0000	0.7879	0.0000	0.0121	0.1000	0.1000
SE	0.7610	0.0000	0.0000	0.0390	0.1000	0.1000
MVI	0.8000	0.0000	0.0000	0.0000	0.1000	0.1000
ANSO	0.7603	0.0000	0.0000	0.0397	0.1000	0.1000

The neurons in input and hidden layers vary 1 through 12 and the number of neuron in output layer is taken as 1 so the total number of examined architectures is 144 for each analyzed time series. *WIC* and *AWIC* values of whole examined architectures are calculated. r_1 and r_2 correlation coefficients are computed by using these values. In Table 1, r_1 and r_2 correlation values and best architectures chosen from second test set for *WIC* and *AWIC* criteria are given.

The weights of *AWIC* are determined by maximizing r_1, which is the correlation coefficient between first and second test sets. The third test set is used in order to observe the consistency of *AWIC* for forecasting. r_2 is the correlation coefficient between second and third test sets. It is clearly seen from Table 1 that *AWIC* is more consistent than *WIC* for all the analyzed time series since r_2 values obtained from *AWIC* are greater than those calculated from *WIC*.

In the optimization process, an optimization problem is solved to determine the coefficients values of the weights of *AWIC*. The optimization problem is written as follows:

$$\max_{w} r_1(w_1, w_2, w_3, w_4) \tag{4.1}$$

subject to

$$0 \leq w_i \leq 1, \quad \text{for } i = 1, 2, 3, 4$$

$$\sum_{i=1}^{4} w_i = 0.8$$

where the function r_1 produce a value that is the correlation coefficient value between the first and second test sets for values of variables w_1, w_2, w_3, and w_4 that are coefficients of *AWIC*. All of the coefficients take values from the interval [0, 1] and the sum of those must be equal to 0.8. As mentioned before, both w_5 and w_6 that are the corresponding weights for *AIC* and *BIC*, respectively, are constant values taken as 0.1. Therefore, there is no need to include these variables in the optimization problem. It should be noted that after the optimization problem is solved, sum of the values of all variables w_1, w_2, w_3, w_4, w_5, and w_6 is expected to have 1.

In the optimization process, to find the best values of the weights in the problem (4.1), we used a MATLAB function called "fmincon". The function "fmincon" is used to find a constrained minimum or maximum of a multivariable function of several variables starting with an initial estimate. It is a built-in function in MATLAB and the detailed information about it can be easily found in the MATLAB help documents. All initial values for w_1, w_2, w_3 and, w_4 are set to 0.2 since Egrioglu et al. used the same values for the weights in *WIC* [12]. Then, the weights of *AWIC* obtained from maximizing r_1 are given in Table 2 for the all time series.

It is observed that the weights of criteria change depending on the analyzed time series.

5. Conclusion

Egrioglu et al. [12] proposed *WIC* in order to select the best architecture for forecasting. The motivation behind the creation of *WIC* is combining the different characteristics of different criteria. *WIC* consists of aggregating the weighted different criteria. These weights are determined intuitively. As a result, it is shown that *WIC* is more consistent that *RMSE*, which is most preferred performance criterion.

In this study, *WIC* is improved by using optimization for determining the weights of criteria. This means that the weights used in *WIC* are determined systematically instead of determining institutively. The newly constructed *WIC* is called *AWIC*. Determining the weights systematically is one of the advantages of *AWIC*. Determining weights based on analyzed time series is second advantage. Finally, more consistency comes with *AWIC*.

In order to show the applicability of *AWIC*, four real time series are analyzed to obtain the best forecasts. The results obtained from employing *AWIC* are compared with those obtained from *WIC*. It is observed that *AWIC* is a more consistent criterion than *WIC*.

References

[1] A. Roy, L.S. Kim, S. Mukhopadhyay, A Polynomial time algorithm for the construction and training of a class of multilayer perceptrons, Neural Networks 6 (1993) 535–545.
[2] Z. Wang, C.D. Massimo, M.T. Tham, A.J. Morris, A procedure for determining the topology of multilayer feedforward neural networks, Neural Networks 7 (1994) 291–300.
[3] N. Murata, S. Yoshizawa, S. Amari, Network information criterion—Determining the number of hidden units for an artificial neural network model, IEEE Transactions on Neural Networks 5 (1994) 865–872.
[4] T.F. Rathbun, S.K. Rogers, M.P. DeSimio, M.E. Oxley, MLP iterative construction algorithm, Neurocomputing 17 (3–4) (1997) 195–216.
[5] J.J.T. Lahnajarvi, M.I. Lehtokangas, J.P.P. Saarinen, Evaluation of constructive neural networks with cascaded Architectures, Neurocomputing 48 (2002) 573–607.
[6] R. Reed, Pruning algorithms a survey, IEEE Transactions on Neural Networks 4 (1993) 740–747.
[7] J. Siestema, R. Dow, Neural net pruning—Why and how? Proceedings of the IEEE International Conference on Neural Networks 1 (1988) 325–333.
[8] S. Buhamra, N. Smaoui, M. Gabr, The Box–Jenkins analysis and neural networks: And time series modeling, Applied Mathematical Modelling 27 (2003) 805–815.
[9] H.C. Yuan, F.L. Xiong, X.Y. Huai, A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy, Computers and Electronics in Agriculture 40 (2003) 57–64.
[10] M. Dam, D.N. Saraf, Design of neural networks using genetic algorithm for on-line property estimation of crude fractionator products, Computers & Chemical Engineering 30 (2006) 722–729.
[11] J. Zeng, H. Guo, Y. Hu, Artificial neural network model for identifying taxi gross emitter from remote sensing data of vehicle emission, Journal of Environmental Sciences 19 (2007) 427–431.
[12] E. Egrioglu, C.H. Aladag, S. Günay, A new model selection strategy in artificial neural networks, Applied Mathematics and Computation 195 (2008) 591–597.
[13] B. Durbin, S. Dudoit, M.J. Van Der Laan, A deletion/substitution/addition algorithm for lassification neural networks, with applications to biomedical data, Journal of Statistical Planning and Inference 138 (2008) 464–488.